



## Circular reasoning rather than cyclic expression

Jensen, Lars Juhl; de Lichtenberg, Ulrik; Jensen, Thomas Skøt; Brunak, Søren; Bork, Peer

*Published in:*  
Genome Biology

*Link to article, DOI:*  
[10.1186/gb-2008-9-6-403](https://doi.org/10.1186/gb-2008-9-6-403)

*Publication date:*  
2008

*Document Version*  
Publisher's PDF, also known as Version of record

[Link back to DTU Orbit](#)

*Citation (APA):*  
Jensen, L. J., de Lichtenberg, U., Jensen, T. S., Brunak, S., & Bork, P. (2008). Circular reasoning rather than cyclic expression. *Genome Biology*, 9(6), 403. <https://doi.org/10.1186/gb-2008-9-6-403>

---

### General rights

Copyright and moral rights for the publications made accessible in the public portal are retained by the authors and/or other copyright owners and it is a condition of accessing publications that users recognise and abide by the legal requirements associated with these rights.

- Users may download and print one copy of any publication from the public portal for the purpose of private study or research.
- You may not further distribute the material or use it for any profit-making activity or commercial gain
- You may freely distribute the URL identifying the publication in the public portal

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Correspondence

## Circular reasoning rather than cyclic expression

Lars Juhl Jensen<sup>\*†‡</sup>, Ulrik de Lichtenberg<sup>‡‡</sup>, Thomas Skøt Jensen<sup>§</sup>,  
Søren Brunak<sup>†¶</sup> and Peer Bork<sup>\*‡</sup>

A response to **Combined analysis reveals a core set of cycling genes** by Y Lu, S Mahony, PV Benos, R Rosenfeld, I Simon, LL Breeden and Z Bar-Joseph. *Genome Biol* 2007, **8**:R146.

Addresses: \*European Molecular Biology Laboratory, D-69117 Heidelberg, Germany. †Novo Nordisk Foundation Center for Protein Research, University of Copenhagen, DK-2200 Copenhagen N, Denmark. ‡LEO Pharma, DK-2750 Ballerup, Denmark. §Center for Biological Sequence Analysis, Technical University of Denmark, DK-2800 Lyngby, Denmark. ¶Max-Delbrück-Centre for Molecular Medicine, D-13092 Berlin, Germany.

<sup>‡</sup>These authors contributed equally to this work.

Correspondence: Peer Bork. Email: bork@embl.de

Published: 23 June 2008

*Genome Biology* 2008, **9**:403 (doi:10.1186/gb-2008-9-6-403)

The electronic version of this article is the complete one and can be found online at <http://genomebiology.com/2008/9/6/403>

© 2008 BioMed Central Ltd

Transcriptome analyses have identified hundreds of genes that are periodically expressed during the mitotic cell cycle in each of four distantly related eukaryotes (budding yeast [1-3], fission yeast [4-6], human [7] and *Arabidopsis thaliana* [8]). In a paper published in *Genome Biology*, Lu and co-workers [9] challenge the results of earlier comparative studies [4,10-15] by claiming that cell-cycle-regulated transcription is much more conserved at the level of individual genes than previously thought. However, we question the validity of their analysis as it relies on circular reasoning, allows evidence from homologous genes to overrule experimental evidence from a gene itself, assesses conservation on the basis of homology rather than orthology, and equates cell-cycle function with cell-cycle regulation.

### Why is the argument circular?

Previous global studies of cell-cycle-regulated expression analyzed the microarray data from each organism individually and then used orthology

relationships derived from sequence homology to compare the regulation of conserved genes. By contrast, Lu and co-workers also use sequence homology to transfer the evidence for periodic expression between sequence homologs within and between organisms [9,16]. If a conserved gene appears periodic in, say, the two yeasts and the plant, then the algorithm may transfer this evidence to the human ortholog of the gene and conclude that it too is periodically expressed. A simplified interpretation of the method is thus that it averages the evidence for and against periodic expression across homologous genes. However, homology transfer is only valid if the transferred property is indeed highly conserved, and it logically follows that one cannot use a method that transfers a property to assess how conserved the property is. The main conclusion of Lu *et al.* [9], namely that cell-cycle regulation is more conserved than suggested by earlier studies, is thus based on circular reasoning as it is a built-in assumption of their method.

Nonetheless, Lu *et al.* say that only “5% to 7% of cycling genes in each of four species have cycling homologs in all other species” and thus agree with previous studies that the vast majority of the cycling genes in an organism do not have cycling homologs in other eukaryotes. When taking into account the limited sensitivity of microarray experiments, we estimate on the basis of our genome-wide comparison that 2% to 8% of the genes in an organism (5 to 22 orthologous groups) belong to the core set of conserved cycling genes (see Supplementary Information of our earlier paper [14]). Whether this is much or little is clearly in the eye of the beholder.

### On which genes do we disagree?

Although the argument for conserved cell-cycle regulation is circular, many of the genes that Lu and co-workers identify as cycling could still be correct. Their method could be useful for upgrading borderline cases, for example, where bad microarray probes give a weak signal for a gene in one of the

organisms. We therefore investigated the disagreements between the lists of periodically expressed genes that arise from the analysis by Lu *et al.* and from our analysis [13,14,17]. Some of the genes on which we disagree are indeed close to the threshold. There are, however, also many cases where the assessment of periodic expression by Lu *et al.* seems completely off. Figure 1 of this Correspondence displays the expression profiles of six such genes. The upper two rows show the data for two budding-yeast kinase genes, *CDC5* and *DBF2*, and their fission-yeast orthologs, *plp1* and *sid2*, all of which have known functions in the cell cycle and have been demonstrated by small-scale experiments to be periodically expressed [13]. Despite consistent periodicity across all five and ten microarray experiments performed on budding and fission yeast, respectively, the analysis by Lu *et al.* [9] shows neither of these genes to be conserved cycling genes.

The opposite scenario is illustrated by the genes *mcm3* and *mcm5*, both of which are mentioned specifically by Lu *et al.* [9] and are even included on the list of fission-yeast genes whose periodicity is supposedly conserved across all four organisms (a class designated by Lu *et al.* as CCC4). These genes exhibit only low-amplitude oscillations in one of ten timecourses, and this is unlikely to be due to active regulation [13]. In fact, *mcm5* is among the 30% least cycling genes in fission yeast according to our analysis [13,14,17]. The combined algorithm by Lu and co-workers thus produces both false negatives and false positives by letting evidence transferred by sequence homology overrule experimental data on the gene itself.

### Are the “conserved cycling genes” orthologous?

Fission yeast *mcm3* and *mcm5* belong to a group of six genes, each encoding a distinct subunit of the hexameric MCM complex, which is involved in initiation of DNA replication. The MCM genes are

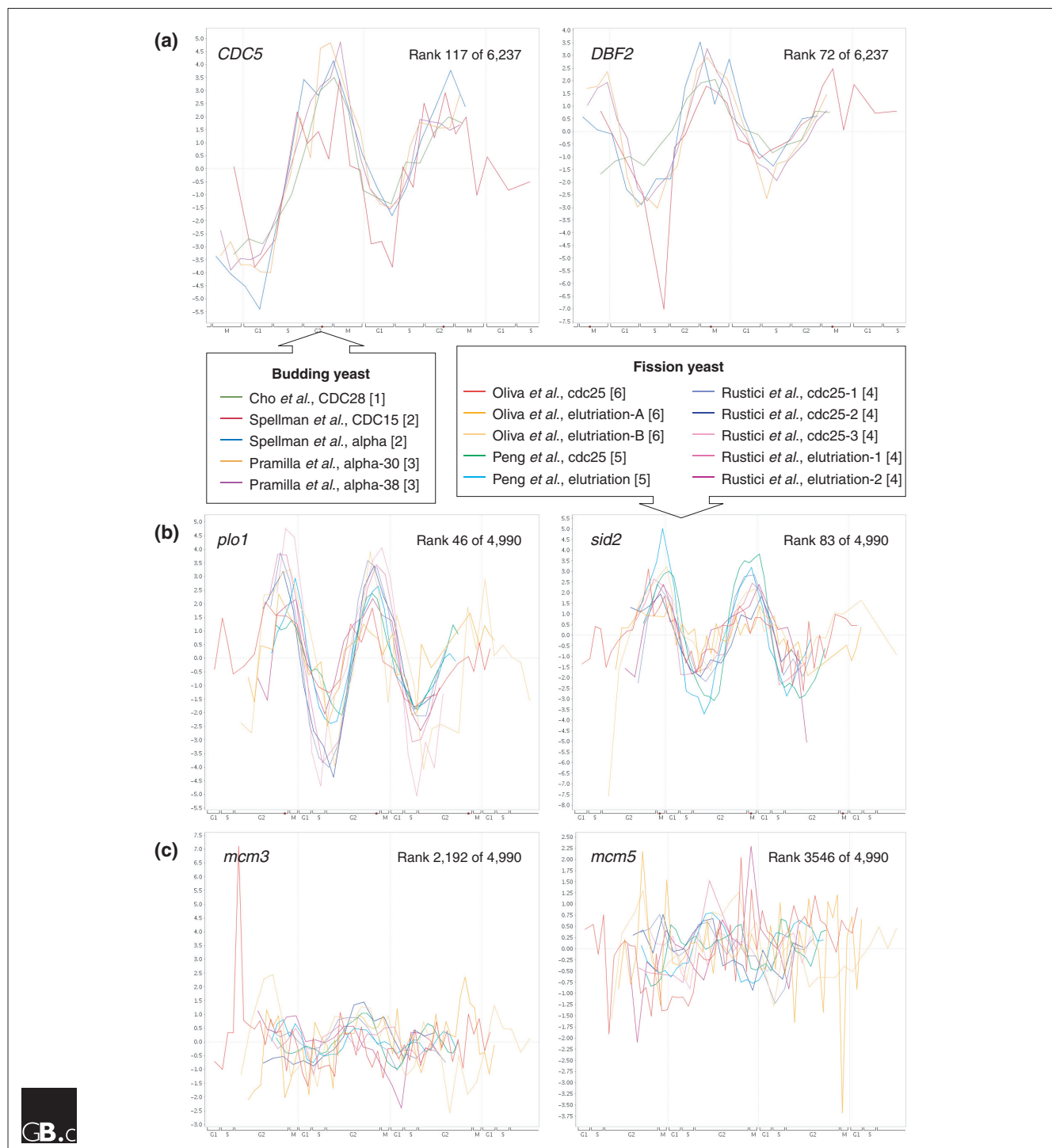
all conserved as 1:1:1:1 orthologs across the four organisms studied [14,18,19]. However, although Lu *et al.* have all six MCM genes from budding yeast as “conserved cycling genes” (CCC4), only *mcm5* is present on all four CCC4 lists. The underlying problem is that their algorithm [9,16], unlike earlier global analyses [4,11,13,14], does not distinguish between orthologs and homologs. A gene cluster may thus contain paralogous genes that arose from gene duplication before the last common ancestor of present-day eukaryotes. This is well illustrated in Figure 1d of [9], in which the four orthologous *CDC6* genes form a cluster that also contains *ORC1* from human and budding yeast (but not from fission yeast and *A. thaliana*). Although both *CDC6* and *ORC1* are presumed to share ancestry with archeal *cdc6* [19,20], they perform distinct, conserved functions in eukaryotes [21]. We consider it questionable to make inferences about, for example, the expression of human *ORC1* based on expression data from budding yeast *CDC6*.

The orthology problem affects many proteins, including probably the most studied of all cell-cycle proteins, the cyclins (Figure 1c in Lu *et al.* [9]). Whereas we agree that the periodic expression of B-type cyclins is conserved [14], the list of human conserved cycling genes from Lu *et al.* also includes those encoding A-, E- and F-type cyclins, although these do not exist in yeasts [18,19]. Tubulins are also listed as conserved cycling genes for each of the four organisms, but the cycling tubulins listed for *A. thaliana* are beta-tubulins, whereas none of the human beta-tubulins cycles. It logically follows that none of the tubulins has periodically expressed orthologs in all four organisms. Systematic, manual checking of all genes on the CCC4 lists reveals that the orthology problem affects almost half of them. The use of the term “conserved cycling gene” is, in our view, therefore misleading, as it does not imply that cyclic expression is conserved between functionally equivalent, orthologous genes.

### Does function imply regulation?

Given the problems described above, how then can it be that the numerous comparisons with other data presented by Lu and co-workers all point in the direction that their lists are better than existing ones? The answer lies in the subtle but important distinction between ‘cell-cycle function’ and ‘cell-cycle regulation’. Figure 1 of this Correspondence exemplifies the difference: whereas all six genes are involved in the cell cycle, only four of them (*Plo1*, *CDC5*, *Sid2*, and *DBF2*) are transcriptionally regulated during the cell cycle. Many of the tests performed by Lu *et al.* to support the validity of their proposed cycling genes do not assess cycling expression *per se*. Datasets from conditions such as stationary-phase budding yeast, nonproliferating human tissues, developmentally arrested *A. thaliana* and nitrogen-starved fission yeast are measures of downregulation in non-proliferating cells, which do not necessarily correlate with cyclic expression. The problem is that any gene involved in the cell cycle should be down-regulated under these conditions - whether it is expressed in a phase-specific manner or not. The authors also analyze the enrichment for essential genes and genes annotated with relevant Gene Ontology terms; however, no statistical analysis can change the fact that these are inherently related to the phenotype or function of a gene rather than to its regulation. The vast majority of the comparisons by Lu *et al.* only show that their set of conserved cycling genes is enriched for genes with cell-cycle function, but not that they are subject to transcriptional cell-cycle regulation. Indeed, we have previously observed that methods with good performance on a benchmark set based on functional evidence often perform poorly on more reliable benchmark sets based on regulatory evidence [22].

Lu and co-workers [9] also compared their list of cycling genes from budding yeast with the targets of nine cell-cycle transcription factors [23,24]. This is, in our view, a much better gold standard as it is based on experimental evidence

**Figure 1**

Expression profiles of six yeast genes across multiple cell-cycle microarray time courses. Expression profiles for **(a)** budding yeast *CDC5* and *DBF2*, and **(b)** their fission yeast orthologs *plo1* and *sid2*. These four genes are all periodically expressed according to our analysis [13,14] but not according to that of Lu and co-workers [9]. **(c)** Conversely, fission yeast *mcm3* and *mcm5* are both periodically expressed according to the analysis of Lu *et al.* [9] but not according to us [13,14,17]. The information in the panels refers to the experiments from which the data come and the method of cell-cycle arrest; for example 'Cho *et al.* [1] CDC28' indicates a time-course experiment in which the cells were arrested using a CDC28 mutant. The values on the y-axis on each profile indicate the log<sub>2</sub> ratio between the expression at a given time point compared with the average expression across the profile. The rank scores show that *plo1* and *sid2* are both among the top 100 cycling genes according to our analysis, whereas *mcm3* and *mcm5* are among the 3,000 least cycling genes. All plots were obtained from the Cyclebase.org database where further details on the normalization procedure and the scoring scheme can also be found [17,38].

that is directly linked to cell-cycle regulation and not to cell-cycle function. However, this benchmark showed that the list proposed by Lu *et al.* [9] and the original list proposed by Spellman *et al.* [2] are equally enriched for targets of cell-cycle transcription factors. Similar benchmarks based on regulatory evidence from the three other organisms even suggest that transfer of evidence between homologous genes leads to a decrease in performance [17].

In summary, homology-based transfer of expression data and other experimental evidence is a powerful strategy for function prediction [25], as protein function is often conserved over long evolutionary distances [20]. However, several studies have shown that the regulation of genes and proteins changes much more rapidly during evolution than their function [4,10-14,26-32]. We have previously shown that, despite the lack of conserved regulation at the single-gene level, the organisms regulate the same protein complexes, but do so via different subunits [14,15]. By transferring cell-cycle expression data between distantly related genes, Lu *et al.* were thus able to identify genes with cell-cycle function that cannot be identified as such on the basis of the expression of the genes themselves (for example, fission yeast *mcm3* and *mcm5*; Figure 1). Selecting the correct evolutionary timescale for the property in question - be that function or regulation - is the key to success for any homology-based method.

Yong Lu, Shaun Mahony, Panayiotis V Benos, Roni Rosenfeld, Itamar Simon, Linda L Breeden and Ziv Bar-Joseph respond:

Despite claims to the contrary from Jensen *et al.*, previous analyses of cell-cycle expression data resulted in opposing views regarding the conservation of expression between different species. While some investigators have concluded that this conservation is surprisingly low [4,14], others have determined that it is rather large. For example, Oliva *et al.* [6] found that

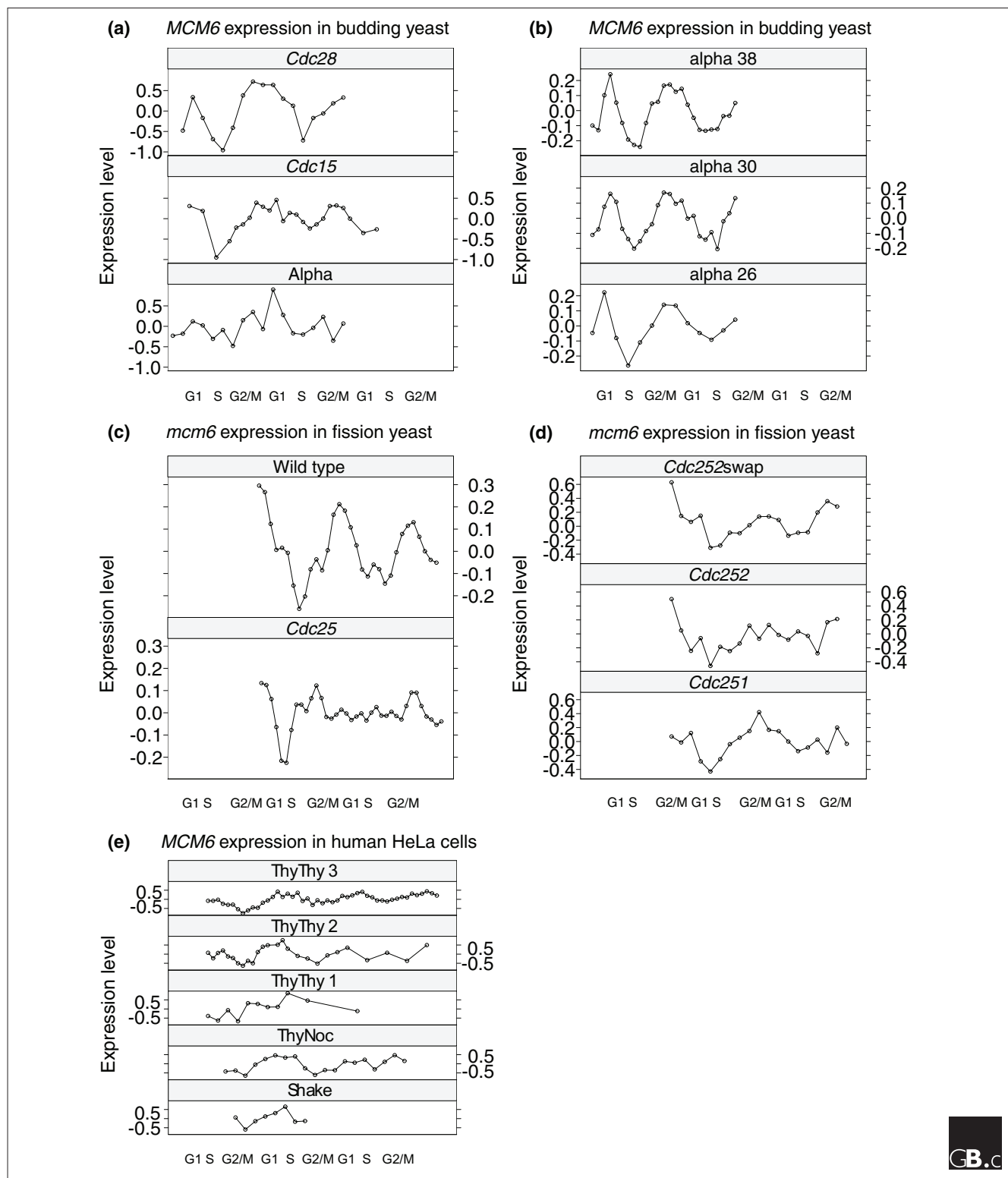
more than 30% of top cycling genes in budding and fission yeast are cycling and conserved in both species, and Ota *et al.* [10] identified more than 15% of cycling human genes as cycling and conserved in plants and yeast. The major reason for this discrepancy seems to be the use of strict thresholding for determining whether a gene is cycling or not. Such an analysis on a species-by-species basis may lead to inconsistencies in cell-cycle assignments. Figure 2 of this Correspondence exemplifies this difficulty. While only expression of the human *Mcm6* gene was determined to be cycling by Jensen *et al.* [14], as Figure 2 shows, its curated homologs in budding and fission yeast (which were annotated as non-cycling by Jensen *et al.*) actually display strong cyclic expression patterns. This is a general problem with cell-cycle analysis. As Figure 3 shows, while some orthologs of cycling budding-yeast genes may fall just below the fission-yeast threshold, they are still (at least weakly) cycling, significantly more than expected by chance, indicating that expression is conserved at a stronger rate than the rate determined by thresholding. To address these issues, we have developed a new method for combining expression data from multiple species [9]. Using our method we concluded that cell-cycle expression is conserved at much higher rates than those claimed by Jensen *et al.* [14].

The central claim Jensen *et al.* raise in this Correspondence is that our method is circular. We believe that they confuse assumptions with circularity. Any computational method relies on specific assumptions and, if these assumptions are wrong, the conclusions of that method may be wrong as well. For example, sequence alignment relies heavily on assumptions regarding the parameters used for match, mismatch and gaps. As Dewey *et al.* [33] nicely show, these parameters can have a big impact on the results of aligning non-coding regions. Nonetheless, researchers have been using these methods for a long time with specific parameter choices and have arrived at very specific

biological conclusions. Like our method, these findings are dependent, at least in part, on the choice of parameters for matches that are directly related to the conclusions drawn. Yet they have proved both useful and accurate when validating with independent data.

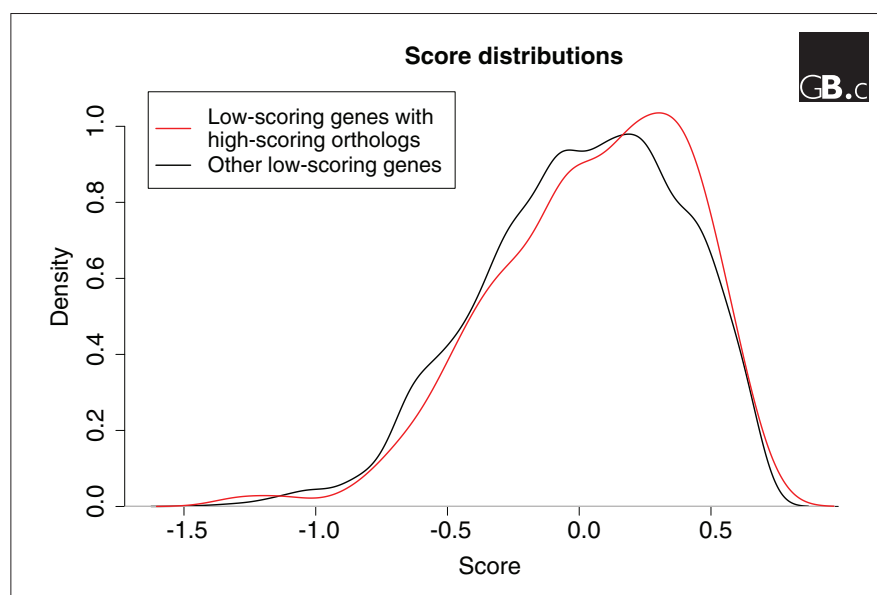
This is exactly the case for our method. It does not rely on circular logic; rather, it uses very specific and widely accepted assumptions. We assume that if two genes have very similar sequence, it increases the likelihood that they perform a similar function. This is the assumption researchers make when using BLAST. When applied to our problem this translates to increased likelihood that genes with a similar sequence share similar cyclic status (either cycling or non-cycling). Note that this assumption is not binding and is only secondary to the actual observed expression values, as we show in Figure 4. Still, as with any other method, we need to decouple our results from our assumptions to demonstrate that our findings are indeed correct. We highlight below the supporting evidence in which we were very careful to control for sequence similarity.

One of the major difficulties in identifying genes whose cell-cycle-regulated transcription is conserved across evolution is that cell-cycle microarray data are noisy and often contradictory. Jensen *et al.* [14] identified the top 300 periodic transcripts from each of four human datasets and found only 63 transcripts in common to all four. With only a 20% overlap between the most periodic 300 transcripts in four datasets from the same organism, there is little doubt that a comparison across four highly diverged species is problematic. The approach of Jensen *et al.* [14] was to use thresholds that are "more conservative than those originally proposed" and to analyze a smaller, more reliable subset of cyclic transcripts. Our goal was not to exclude, but to capture as many cyclic transcripts as possible, with the view that interesting candidates could be subjected to further verification.

**Figure 2**

Expression values for *MCM6* in humans, budding yeast, and fission yeast. Values are log ratios between synchronized and unsynchronized cells. **(a,b)** Expression profiles of budding yeast *MCM6* under different cell-cycle arrest methods [2,3]. **(c,d)** Expression of fission yeast *mcm6* under different arrest methods [4,5]. **(e)** Expression of *MCM6* in human HeLa cells [7]. Cell-cycle stages are shown underneath each panel. Jensen et al. [14] claim that although human *MCM6* is cycling at the transcriptional level, its homologs in budding yeast and fission yeast do not cycle. As (a-d) show, the expression of yeast *MCM6* seems more cyclic than that of human *MCM6*, highlighting the limitations of species-by-species thresholding.



**Figure 3**

Score distributions for fission-yeast genes that are ranked below the cycling score threshold. The red curve is the distribution of 350 fission-yeast orthologs of cycling budding-yeast genes. The black curve is the distribution of all the other 3,641 fission-yeast genes. Density is the distribution density for each of the different scores. As can be seen, the red curve is highly skewed to the right (higher score). In fact, the difference between the two curves is significant, with a  $p$ -value of 0.01 (Wilcoxon rank-sum test). Thus, while orthologs of cycling budding-yeast genes may fall just below the fission-yeast threshold, they are still at least weakly cycling, much more so than expected by chance, indicating that expression is conserved at a much stronger rate than the rate determined by thresholding-based methods.

Our approach was motivated by the plot in Figure 2, which shows that fission-yeast orthologs of cycling budding-yeast genes fall just below the fission-yeast threshold for periodicity far more than expected from chance ( $p$ -value  $< 0.01$  using Wilcoxon rank-sum test,  $p$ -value  $< 0.03$  using Kolmogorov-Smirnov double-sided test). We have attempted to capture these borderline genes by lowering the threshold for borderline genes if their homologs in other species are cyclic and raising them if they are not cyclic. This strategy will certainly lead to more false assignments, but it has also allowed us to identify hundreds of promising candidates for further investigation. Still, almost all the genes that are elevated to a cyclic status by our method have a rather high cyclic expression score to begin with. Figure 3 shows the difference between the initial score (based on expression alone) and the posterior score from our method. As can be seen in the plot, the ranks for most genes do not change much.

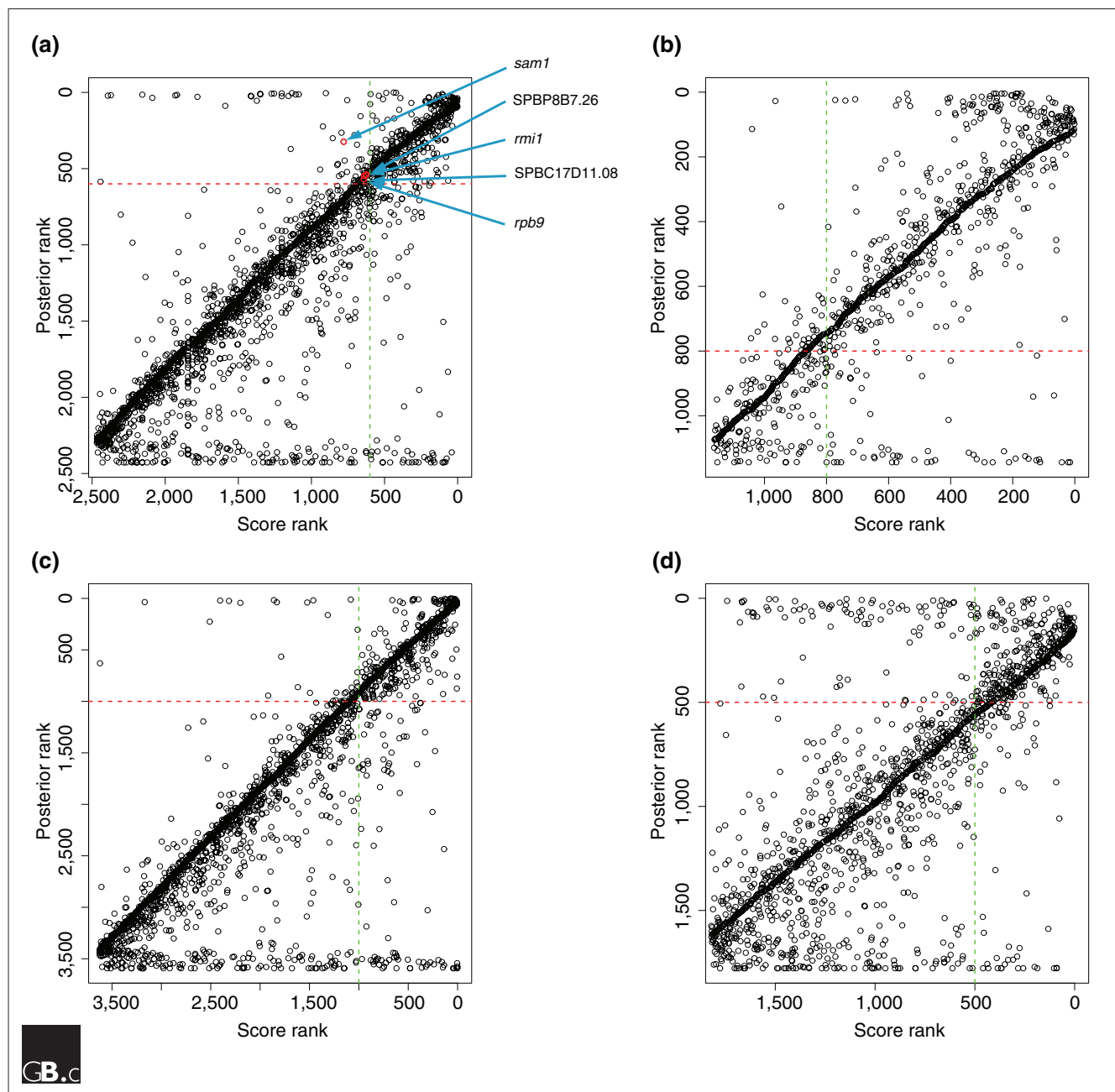
Jensen *et al.* also question the complementary datasets we used to validate the CCC sets identified by our algorithm. They claim that the complementary datasets we used only point to cell-cycle function rather than cell-cycle regulation. However, the 'functional' rather than regulatory identification claim does not provide an explanation as to how our algorithm was able to identify these 'functional' cell-cycle genes. In our analysis we used controls for both types of data (expression and sequence). Specifically, for the essentiality analysis we show that only 16% of cycling yeast genes are essential. If one uses sequence data, so that only genes with conserved homologs in other species are retained, this percentage increases to 27%. If what we find is indeed functional rather than regulatory signal, cyclic expression in other species would not have been a factor and the only advantage we would have come from using sequence data. However, when we use both sequence conservation and conserved cyclic expression, as

determined by our method the percentage rises to 46%, a more than 70% increase over sequence alone. Similar results were obtained for the human conserved set. We have repeated this type of positive control for the other types of complementary analysis and have shown that expression conservation leads to much stronger cell-cycle characteristics.

We have also carried out direct regulatory analysis. Table 1 in our original paper [9] presents the result of motif search methods for genes in CCC2, the set of cycling genes conserved between the two yeasts. We show that these genes have a remarkably well conserved motif for G1 and some of the S-phase transcription factors. In sharp contrast, non-cycling homologs of genes in CCC2 do not have these motifs conserved. The fact that motif conservation agrees with our expression conservation findings is a strong support for the CCC2 set assignment.

The other major issue raised here by Jensen *et al.* relates to the problem of identifying conserved periodic genes whose products carry out the same function in all four of these highly divergent species. Jensen *et al.* [14] used a combination of sequence similarity and manual curation to identify orthologous groups. In most cases, it cannot be determined whether these groups are really functionally equivalent or whether all such groups have been identified. Nevertheless, on the basis of these assignments, only a quarter of all the cycling genes they studied had orthologs in all four species and these form the basis for their comparison. Of the 60 cycling genes in *Arabidopsis* with orthologs in all four species, one-third of their orthologs also cycle in pairwise comparisons with each of the other three species, but only five cycle in all four species. All five of these orthology groups represent well studied genes and nothing new was identified.

We purposely avoided restricting our analysis only to genes with clear

**Figure 4**

Comparison of expression score ranks and posterior ranks. **(a)** The expression score rank and posterior rank for fission-yeast genes. The x-axis is the expression score rank (the lower the rank the more cyclic the gene is determined to be by the scoring method) and the y-axis is the rank based on our method (again, the lower the better). As can be seen, the ranks for most of the genes do not change much. The red dashed line represents the posterior threshold used to select cycling genes, and the green dashed line is the corresponding threshold if only expression scores are used. Almost all genes that are elevated by our method to a cyclic status have a rather high cyclic expression score (though some are not as high as the cutoff for score alone, which is where the two methods differ). Five selected genes are highlighted by red circles. These genes would have been missed if only expression scores were used to determine cyclic status, because their scores would be just below the cutoff. While Jensen *et al.* [14] do not assign cyclic status to these genes, *sam1* was also identified as cycling by Peng *et al.* [5], *SPBC17D11.08* was included in the list by Rustici *et al.* [4], and *rpb9* was identified by both Oliva *et al.* [6] and Peng *et al.* [5]. The other two genes, *SPBP8B7.26* and *rmi1*, are missing from all three studies, even though their profiles appear cyclic (not shown). **(b-d)** Similar plots for (b) budding yeast [2,3], (c) human [7], and (d) *Arabidopsis* [39].

orthologs across species. Rather, we used BLAST analysis followed by a Markov cluster algorithm [34], which

leads to the identification of multi-domain homologous proteins. This difference between the definitions of

homologs impacts on the conclusions reached by us and Jensen *et al.* Our method results in large families that



show high homology overall but cannot be parsed into one-to-one orthologous pairs across species. In our original paper [9], we presented analysis of the results of this procedure for the CCC2 set of conserved cycling genes. We found that 82% of budding yeast genes in CCC2 are indeed curated homologs of the fission yeast CCC2 genes [35], a very high rate that indicates the accuracy of the resulting CCC2 set.

As we compare the genes from more divergent species, we are much less likely to be able to ascribe functional equivalence to any given pair. This is especially true for signaling and regulatory proteins that often arise from duplicated genes, and which cannot be forced into functionally equivalent orthology groups until we have a complete understanding of what they do in every species. Jensen *et al.* are correct that there is no cyclin E ortholog in yeast. There is also no cyclin E in *Arabidopsis* [36]. However, all four species encode related cyclin genes carrying out functions in late G1 that are important for the transition to S phase, and most of these cyclins are cell-cycle-regulated at the transcriptional level. These are the very types of gene products that we are most interested in identifying.

Towards this end we used an objective and comprehensive strategy for identifying multi-domain sequence homologies across all four genomes. In so doing, we have identified groups of genes that share some truly remarkable properties. The 72 conserved cyclic budding-yeast genes that are also conserved in fission yeast and humans (CCC3) are eight times more likely to be targets of cyclin-dependent kinases than those tested at random, and six times more likely to be involved in protein-protein interactions. Some of these genes encode unexpected proteins (for example, alkaline phosphatase and metal transporters) and there are others about which nothing is known. To further study this set we carried out new experiments [37] to identify the set of cycling genes in primary human cells

(our previous analysis as well as that analysis of Jensen *et al.* [14] is based on expression data from transformed (HeLa) cells). As we discuss in [37], the set of genes cycling in primary cells is significantly more enriched than the HeLa set for orthologs of cycling genes in budding and fission yeast. We hope that our study will spur the collection of more cell-cycle data and the development of new strategies for identifying conserved periodically transcribed genes.

Correspondence should be sent to Ziv Bar-Joseph: Department of Computer Science, Carnegie Mellon University, 5000 Forbes Avenue, Pittsburgh, PA 15213, USA. Email: zivbj@cs.cmu.edu

## References

1. Cho RJ, Campbell MJ, Winzeler EA, Steinmetz L, Conway A, Wodicka L, Wolfsberg TG, Gabriellian AE, Landsman D, Lockhart DJ, Davis RW: **A genome-wide transcriptional analysis of the mitotic cell cycle.** *Mol Cell* 1998, **2**:65-73.
2. Spellman PT, Sherlock G, Zhang MQ, Iyer VR, Anders K, Eisen MB, Brown PO, Botstein D, Futcher B: **Comprehensive identification of cell cycle-regulated genes of the yeast *S. cerevisiae* by microarray hybridization.** *Mol Biol Cell* 1998, **9**:3273-3297.
3. Pramila T, Wu W, Miles S, Noble WS, Breeden LL: **The Forkhead transcription factor Hcm1 regulates chromosome segregation genes and fills the S-phase gap in the transcriptional circuitry of the cell cycle.** *Genes Dev* 2006, **20**:2266-2278.
4. Rustici G, Mata J, Kivinen K, Lió P, Penkett CJ, Burns G, Hayles J, Brazma A, Nurse P, Bähler J: **Periodic gene expression program of the fission yeast cell cycle.** *Nat Genet* 2004, **36**:809-817.
5. Peng X, Karuturi RK, Miller LD, Lin K, Jia Y, Kondu P, Wang L, Wong LS, Liu ET, Balasubramanian MK, Liu J: **Identification of cell cycle-regulated genes in fission yeast.** *Mol Biol Cell* 2005, **16**:1026-1042.
6. Oliva A, Rosebrock A, Ferrezuelo F, Pyne S, Chen H, Skiena S, Futcher B, Leatherwood J: **The cell cycle-regulated genes of *Schizosaccharomyces pombe*.** *PLoS Biol* 2005, **3**:e225.
7. Whitfield ML, Sherlock G, Saldanha AJ, Murray JI, Ball CA, Alexander KE, Matese JC, Perou CM, Hurt MM, Brown PO, Botstein D: **Identification of genes periodically expressed in the human cell cycle and their expression in tumors.** *Mol Biol Cell* 2002, **13**:1977-2000.
8. Menges M, Hennig L, Grussem W, Murray JAH: **Genome-wide gene expression in an *Arabidopsis* cell suspension.** *Plant Mol Biol* 2003, **53**:423-442.
9. Lu Y, Mahony S, Benos PV, Rosenfeld R, Simon I, Breeden LL, Bar-Joseph Z: **Combined analysis reveals a core set of cycling genes.** *Genome Biol* 2007, **8**:R146.
10. Ota K, Goto S, Kanehisa M: **Comparative analysis of transcriptional regulation**

**in eukaryotic cell cycles.** In *Proc Fourth Int Workshop Bioinf Systems Biol 2004* Poster Abstracts: 26-27. [http://www.jsbi.org/modules/journal/index.php/IBSB04/IBSB04Poster.html]

11. Sherlock G: **STARTing to recycle.** *Nat Genet* 2004, **36**:795-796.
12. Dyczkowski J, Vingron M: **Comparative analysis of cell cycle regulated genes in eukaryotes.** *Genome Informatics* 2005, **16**:125-131.
13. Marguerat S, Jensen TS, de Lichtenberg U, Wilhelm BT, Jensen LJ, Bähler J: **The more the merrier: comparative analysis of microarray studies on cell cycle-regulated genes in fission yeast.** *Yeast* 2006, **23**:261-277.
14. Jensen LJ, Jensen TS, de Lichtenberg U, Brunak S, Bork P: **Co-evolution of transcriptional and posttranslational cell-cycle regulation.** *Nature* 2006, **443**:594-597.
15. de Lichtenberg U, Jensen TS, Brunak S, Bork P, Jensen LJ: **Evolution of cell-cycle control: same molecular machines, different regulation.** *Cell Cycle* 2007, **6**:1819-1825.
16. Lu Y, Rosenfeld R, Bar-Joseph Z: **Identifying cycling genes by combining sequence homology and expression data.** *Bioinformatics* 2006, **22**:e314-e322.
17. Gauthier NP, Larsen ME, Wernersson R, de Lichtenberg U, Jensen LJ, Brunak S: **Cyclebase.org - a comprehensive multi-organism online database of cell cycle experiments.** *Nucleic Acids Res* 2008, **36**:D854-D859.
18. Li H, Coghlan A, Ruan J, Coin LJ, Hériché JK, Osmotherly L, Li R, Liu T, Zhang Z, Bolund L, Wong GKS, Zheng W, Dehal P, Jun Wang J, Durbin R: **TreeFam: a curated database of phylogenetic trees of animal gene families.** *Nucleic Acids Res* 2006, **34**:D572-D580.
19. Tatusov RL, Fedorova ND, Jackson JD, Jacobs AR, Kiryutin B, Koonin EV, Krylov DM, Mazumder R, Mekhedov SL, Nikolskaya AN, Rao BS, Smirnov S, Sverdlov AV, Vasudevan S, Wolf YI, Yin JJ, Natale DA: **The COG database: an updated version includes eukaryotes.** *BMC Bioinformatics* 2003, **4**:41.
20. Tatusov RL, Koonin EV, Lipman DJ: **A genomic perspective on protein families.** *Science* 1997, **278**:631-637.
21. Bell SP, Dutta A: **DNA replication in eukaryotic cells.** *Annu Rev Biochem* 2002, **71**:333-374.
22. de Lichtenberg U, Jensen LJ, Fausbøll A, Jensen TS, Bork P, Brunak S: **Comparison of computational methods for the identification of cell cycle regulated genes.** *Bioinformatics* 2005, **21**:1164-1171.
23. Simon I, Barnett J, Hannett N, Harbison CT, Rinaldi NJ, Volkert TL, Wyrick JJ, Zeitlinger J, Gifford DK, Jaakkola TS, Young RA: **Serial regulation of transcriptional regulators in the yeast cell cycle.** *Cell* 2001, **106**:697-708.
24. Lee I, Date SV, Adai AT, Marcotte EM: **A probabilistic network of yeast genes.** *Science* 2004, **306**:1555-1558.
25. Bork P, Jensen LJ, von Mering C, Ramani AK, Lee I, Marcotte EM: **Protein interaction networks from yeast to human.** *Curr Opin Struct Biol* 2004, **14**:292-299.
26. Ihmels J, Bergmann S, Berman J, Barkai N: **Comparative gene expression analysis by a differential clustering approach: application**

- to the *Candida albicans* transcription program. *PLoS Genet* 2005, **1**:e39.
27. Ihmels J, Bergmann S, Berman J, Barkai N: **Rewiring of the yeast transcriptional network through the evolution of motif usage.** *Science* 2005, **309**:938-940.
  28. Marino-Ramirez L, Jordan IK, Landsman D: **Multiple independent evolutionary solutions to core histone gene regulation.** *Genome Biol* 2006, **7**:R122.
  29. Tirosh I, Weinberger A, Carmi M, Barkai N: **A genetic signature of interspecies variations in gene expression.** *Nat Genet* 2006, **38**:830-834.
  30. Borneman AR, Gianoulis TA, Zhang ZD, Yu H, Rozowsky J, Seringhaus MR, Wang LY, Gerstein M, Snyder M: **Divergence of transcription factor binding sites across related yeast species.** *Science* 2007, **317**:815-819.
  31. Odom DT, Dowell RD, Jacobsen ES, Gordon W, Danford TW, MacIsaac KD, Rolfe PA, Conboy CM, Gifford DK, Fraenkel E: **Tissue-specific transcriptional regulation has diverged significantly between human and mouse.** *Nat Genet* 2007, **39**:730-732.
  32. Tan T, Shlomi T, Feizi H, Ideker T, Sharan R: **Transcriptional regulation of protein complexes within and across species.** *Proc Natl Acad Sci USA* 2007, **104**:1283-1288.
  33. Dewey CN, Huggins PM, Woods K, Sturm-fels B, Pachter L: **Parametric alignment of *Drosophila* genomes.** *PLoS Comp Biol* 2006, **2**:e73.
  34. Enright AJ, Van Dongen S, Ouzounis CA: **An efficient algorithm for large-scale detection of protein families.** *Nucleic Acids Res* 2002, **30**:1575-1584.
  35. Penkett CJ, Morris JA, Wood V, Bähler J: **YOGY: a web-based, integrated database to retrieve protein orthologs and associated Gene Ontology terms.** *Nucleic Acids Res* 2006, **34**:W330-W334.
  36. Menges M, de Jager SM, Gruissem W, Murray JA: **Global analysis of the core cell cycle regulators of *Arabidopsis* identifies novel genes, reveals multiple and highly specific profiles of expression and provides a coherent model for plant cell cycle control.** *Plant J* 2005, **41**:546-566.
  37. Bar-Joseph Z, Siegfried Z, Brandeis M, Brors B, Lu Y, Eils R, Dynlacht BD, Simon I: **Genome-wide transcriptional analysis of the human cell cycle identifies genes differentially regulated in normal and cancer cells.** *Proc Natl Acad Sci USA* 2008, **105**:956-961.
  38. Cyclebase [<http://cyclebase.org>]
  39. Menges M, Hennig L, Gruissem W, Murray JA: **Cell cycle-regulated gene expression in *Arabidopsis*.** *J Biol Chem* 2002, **277**:41987-42002.